Senkbeil, Martin; Ihme, Jan Marten; Wittwer, Jörg

The Test of Technological and Information Literacy (TILT) in the National Educational Panel Study: development, empirical testing, and evidence for validity

Journal for educational research online 5 (2013) 2, S. 139-161

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 5 (2013), No. 2, 139–161

Martin Senkbeil, Jan Marten Ihme & Jörg Wittwer

The Test of Technological and Information Literacy (TILT) in the National Educational Panel Study: Development, empirical testing, and evidence for validity

Abstract

In this article, we present an assessment framework for assessing information and communication technologies literacy (ICT literacy) in the context of the National Educational Panel Study (NEPS). NEPS is the first study aiming to examine ICT literacy longitudinally across the lifespan using objective paper-pencil test instruments. To do so, it is necessary to develop reliable and valid test instruments that are capable of assessing ICT literacy longitudinally. On the basis of an assessment framework, we developed in an initial step an item pool for assessing the ICT literacy of secondary school students and tested the item pool in three pilot studies (Grade 5, 7 and 9). The item and scale analyses provided satisfying results and the item pool proved to be suitable for assessing ICT literacy longitudinally. In addition, validity analyses showed a satisfactory content validity and a high test fairness with regard to gender and school type. Concerning criterion validity, there were expected correlations with variables of computer familiarity.

Keywords

ICT literacy; Competence-based test; Validity

Dr. Martin Senkbeil (corresponding author) · Dr. Jan Marten Ihme, Leibniz Institute for Science and Mathematics Education (IPN) at the University of Kiel, Olshausenstraße 62, 24118 Kiel, Germany
e-mail: senkbeil@ipn.uni-kiel.de
ihme@ipn.uni-kiel.de

Prof. Dr. Jörg Wittwer, Department of Empirical Research on Instruction and Learning, Educational Institute, University of Göttingen, Waldweg 26, 37075 Göttingen, Germany
e-mail: joerg.wittwer@sowi.uni-goettingen.de

Martin Senkbeil, Jan Marten Ihme & Jörg Wittwer

Test zur Erfassung technologischer und informationsbezogener Literacy (TILT) im Nationalen Bildungspanel: Entwicklung, empirische Überprüfung und Validitätshinweise

Zusammenfassung

Im Mittelpunkt dieses Beitrags steht die Vorstellung und empirische Erprobung einer Rahmen- und Testkonzeption zur Erfassung von Information and Communication Technologies Literacy (ICT Literacy) im Rahmen der National Educational Panel Study (NEPS). Das Ziel dieser Studie ist es, ICT Literacy anhand längsschnittlich angelegter Datenerhebungen über die gesamte Lebensspanne und mit Hilfe objektiver und zunächst Papier-und-Bleistift-basierter Testverfahren zu erfassen. Deshalb ist es notwendig, reliable und valide Testinstrumente zu entwickeln, die sich für die längsschnittliche Erfassung von ICT Literacy eignen. Hierfür wurde auf der Grundlage der Rahmen- und Testkonzeption zunächst ein Itempool für Schülerinnen und Schüler der Sekundarstufe I entwickelt und in drei Pilotstudien (Klassenstufen 5, 7 und 9) erprobt. Für den Itempool wurden zufriedenstellende Item- und Skalenwerte ermittelt. Die Ergebnisse wiesen darüber hinaus auf die Eignung des Instrumentariums für längsschnittliche Erhebungen hin. Die Validitätsprüfungen sprachen für eine ausreichende inhaltliche Validität des Itempools und seine Fairness im Hinblick auf Geschlecht und Schulart. Im Sinne der Kriteriumsvalidität ergaben sich inhaltlich gut interpretierbare Zusammenhänge mit Merkmalen der Computervertrautheit.

Schlagworte

ICT Literacy; Kompetenzbasierter Test; Validität

1. Introduction

The ability to engage effectively with information and communication technologies not only plays an important role in many workplace settings, but is also becoming increasingly important in people's everyday lives (Educational Testing Service [ETS], 2002; Partnership of 21st Century Skills, 2007). Thus, literacy in using information and communication technologies (ICT literacy) is a key social and cultural proficiency necessary for successful participation in society and for achieving personal and professional goals (e.g., Blossfeld, 2010; Kozma, 2009). More recent conceptualizations of ICT literacy (e.g., ETS, 2002; Fraillon & Ainley, 2010; MCEETYA, 2007) are not exclusively confined to technological literacy, that is, functional knowledge of hardware and software applications (see, e.g., Markauskaite, 2006). Instead, information literacy, that is, the ability to use digital media to locate and critically evaluate information and to use it effectively for one's own purposes, also plays an important role (Pinto, Cordon, & Diaz, 2010). Information literacy, for example, requires problem-solving skills, metacognition, and critical thinking (e.g., ETS, 2002; Partnership for 21st Century Skills, 2007).

Against this background, ICT literacy is assumed to function as a meta-competence that helps people to acquire other important competencies and skills that are relevant in educational and work settings across the lifespan (Partnership for 21st Century Skills, 2007, p. 10–11; see also Blossfeld, von Maurice, & Schneider, 2011; Kozma, 2009). Since globalization processes require more flexibility and adaptability both at work and in society, the ability to acquire new knowledge and new skills in a self-regulated manner and with the help of information and communication technologies has become an important precondition for both finding new jobs and acting as responsible citizens (Blossfeld, von Maurice, et al., 2011, p. 7). Accordingly, deficits in ICT literacy are expected to lead to social disparities (e.g., van Dijk, 2006). For example, it is conjectured that in early phases of life (e.g., in primary school), differences in the social background are responsible for differences in ICT literacy and that these differences increase across the lifespan (e.g., Poynton, 2005). However, there is not much empirical research examining the role of ICT literacy as a meta-competence and its effects on social disparities across the lifespan (e.g., van Dijk, 2006).

The design of the National Educational Panel Study (NEPS) principally allows for studying meta-competencies such as ICT literacy across the lifespan ranging from early kindergarten to later adulthood (for more details, see e.g., Blossfeld, von Maurice, et al., 2011; Blossfeld, Roßbach, & von Maurice, 2011). However, in order to do so, it is necessary to have reliable and valid test instruments that assess ICT literacy longitudinally. Existing test instruments are usually confined to specific age groups (e.g., university students), are not constructed on a theoretical basis (cf. van Deursen & van Dijk, 2010), or assess only specific aspects of ICT literacy such as technical literacy (e.g., Richter, Naumann, & Horz, 2010) or internet literacy (e.g., Hargittai, 2003). Thus, it is necessary to develop new test instruments that are capable of assessing – more broadly and on a theoretical basis – ICT literacy across the lifespan.

Another point warranting consideration in the development of this instrument is that, whereas the first waves of the NEPS assessments are administered as paper-and-pencil tests (PPT), a switch to computer-based testing (CBT), which will also facilitate the use of innovative item formats, is planned for later waves (Kröhne & Martens, 2011). Accordingly, the conceptual framework and the basic item format have to be developed in such a way that they can be implemented in both formats (PPT, CBT) and are compatible with innovative forms of assessment (e.g., interactive simulations).

Therefore, in this article, we present the assessment framework of the test instruments developed to assess ICT literacy in NEPS. In addition, we provide results concerning the psychometric quality and the validity of the test instruments. In an initial step, we developed an item pool for assessing ICT literacy of secondary school students and tested the item pool in three age cohorts (Grade 5, Grade 7, and Grade 9). In doing so, we addressed the following goals:

1) The item pool is tested for its psychometric quality, dimensionality, and appropriateness for assessing ICT literacy longitudinally.
2) The item pool is tested for different aspects of validity: (a) Using expert ratings, the content validity of the item pool is tested. (b) Analyses of differential item functioning (DIF) are undertaken to check for the test fairness of the item pool. (c) The concurrent validity of the item pool (as part of the criterion validity) is examined by using different background variables (e.g., school track) and variables related to computer usage (e.g., using computers for entertainment).

2. The conceptual and assessment framework for ICT literacy in NEPS

NEPS is funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) and conducted by a consortium under the auspices of the Institute of Longitudinal Educational Research (INBIL) at the University of Bamberg (for more details, see Blossfeld, Roßbach, et al., 2011). The aim of NEPS is to examine domain-specific competencies (e.g., reading, mathematics, science) and meta-competencies (e.g., ICT literacy) from different age cohorts across the entire lifespan ranging from kindergarten to later adulthood. NEPS is a multi-cohort-sequence design in that five cohorts were sampled at the beginning of the study in 2010 (kindergarten, Grade 5, Grade 9, university students, and adults). The competence domains are continuously assessed throughout the lifespan. Therefore, it is possible to examine developmental changes in the competence domains as a function of different educational stages (e.g., kindergarten, primary school, secondary school).

In the context of this study, we conceptualize ICT literacy as a unidimensional construct comprising the facets of technological and information literacy (e.g., ETS, 2002; Fraillon & Ainley, 2010; MCEECDYA, 2010; MCEETYA, 2007). In line with the literacy concepts of large-scale international assessments (e.g., the Programme for International Student Assessment, PISA), we define ICT literacy from a functional perspective in the context of NEPS. That is, functional literacy is understood to include the knowledge and skills that people need to lead satisfying lives in personal and economic terms in modern-day societies. This leads to an assessment framework that relies heavily on everyday problems which are more or less distant to school curricula (e.g., using the internet for shopping; see Weinert et al., 2011).

As a basis for the construction of an instrument that assesses ICT literacy in NEPS, we use the framework "Digital transformation. A framework for ICT literacy" of the International ICT Literacy Panel (ETS, 2002). The framework already resulted in the development of the ETS iSkills, an assessment of ICT literacy (Katz, 2007), and the International Computer and Information Literacy Study (ICILS;

Fraillon & Ainley, 2010). The framework identifies seven process components of ICT literacy that represent the knowledge and skills needed for a problem-oriented use of modern ICT (ETS, 2002; Katz, 2007). The seven process components are defined as follows:

1) *Define*: basic knowledge of hardware components, operating systems, and relevant software applications (e.g., word processing, e-mail);
2) *Access*: knowledge of basic operations used to retrieve information (e.g., entering a search term in an internet browser, opening and saving a document);
3) *Manage*: the ability to find information within a program (e.g., retrieving information from tables, processing the hits returned by a search engine);
4) *Create*: the ability to create and edit documents and files (e.g., setting up tables, creating formulas);
5) *Integrate*: the ability to retrieve information efficiently (e.g., entering appropriate search terms), to compare it in terms of specific criteria (e.g., sorting datasets), or to organize and structure information with the aid of software applications (e.g., presenting the information retrieved in a table);
6) *Evaluate*: the ability to assess information and to use it as the basis for informed decisions (e.g., assessing the credibility of the information retrieved);
7) *Communicate*: the ability to communicate information in appropriate and understandable form (e.g., writing e-mails, creating meaningful charts).

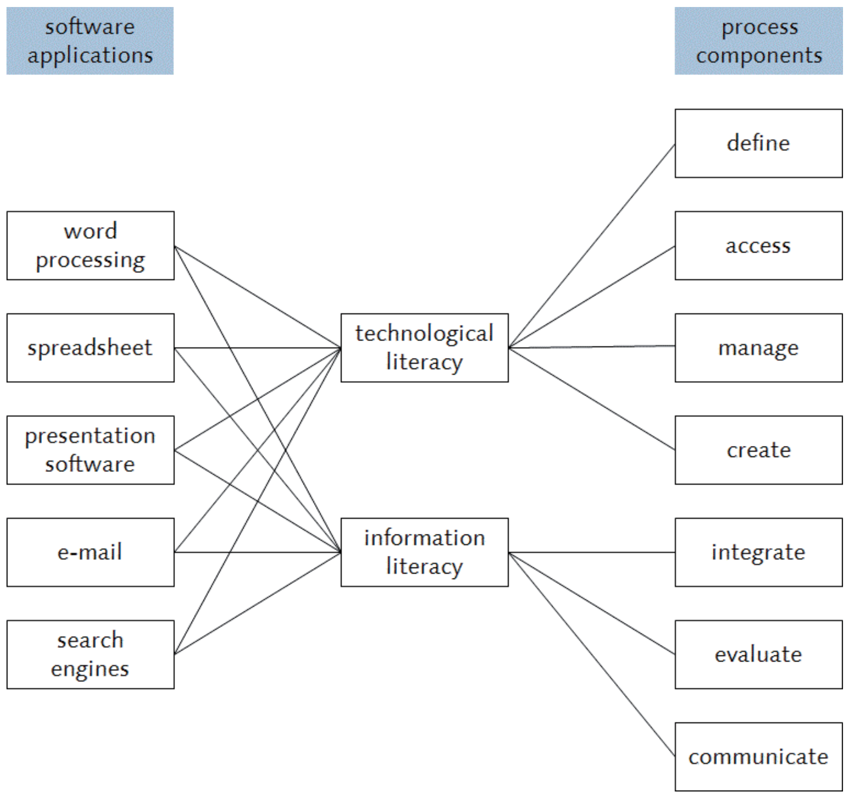
On the basis of the framework, we develop the Technological and Information Literacy Test instrument (TILT) of ICT literacy in NEPS, consisting of one test for each cohort. The first four process components (*Define*, *Access*, *Manage*, *Create*) refer to the facet of technological literacy whereas the latter three process components (*Integrate*, *Evaluate*, *Communicate*) refer to the facet of information literacy. It is assumed that the knowledge needed to solve tasks related to technological literacy are, for the most part, automated routines that have been acquired through practice and can thus be described as schemata stored in long-term memory (Dutke & Reimer, 2000). The routine tasks (e.g., formatting a document) primarily require declarative knowledge ("knowing what"; e.g., terminological knowledge) and procedural knowledge ("knowing how"; e.g., practical skills in using specific computer applications; Mayer, 2002). In contrast, tasks related to information literacy require not only the retrieval of knowledge from long-term memory but also the generation of new knowledge, for example, through an internet search for relevant information on a topic (see, e.g., van Deursen & van Dijk, 2010). In addition, conceptual knowledge is needed that enables students to recognize and represent classifications, structures, and principles to be applied to the problem situation at hand (Krathwohl, 2002).

Apart from the process components, the test construction of TILT is guided by a categorization of software applications that are used to locate, process, present, and communicate information (see International Society for Technology in Education [ISTE], 1998):

1) word processing and operating systems;
2) spreadsheets;
3) presentation and graphics software;
4) e-mail and other communication applications (e.g., forums);
5) internet-based search engines and databases.

With regard to the content validity of the item pool, we seek to achieve a roughly equal distribution of items across the seven process components and the five software applications. Moreover, we ensure to fill as many cells resulting from crossing the two dimensions *process components* and *software applications*. The resulting conceptual framework provides the basis for the construction of all item pools and tests across the different age groups and educational stages assessed in NEPS (see Figure 1 for an illustration). The difficulty level and contextualization of each item (in a situation as authentic as possible) are adapted to the age cohort under investigation.

Figure 1: Assessment framework for ICT literacy



3. Test material

As the focus of assessing ICT literacy in the first phase of NEPS is on lower secondary students, we first developed a pool of items to assess ICT literacy in this age group. The psychometric quality of this item pool and selected aspects of its validity were tested in three pilot studies with samples of Grade 5, 7, and 9 students (see Table 2). The item pool comprised 140 items in multiple-choice format. Based on the item parameters calculated in pre-tests (i.e., in cognitive labs and testing some school classes), it was possible to administer items that were consistent with a student's predicted ability level in the pilot studies.

The distribution of test items across the two dimensions (see Table 1) shows that the pool of items developed was, to a large extent, consistent with the conceptual principles guiding our test construction. It was only for the process components *Create* and *Evaluate* that the numbers of items were notably above or below average (29 and 9 items, respectively). In all other cases, the items were almost equally distributed across the five software applications. Moreover, almost 80% of the cells in the cross-tabulation matrix included at least one test item (in most cases, several items).

Table 1: Distribution of the number of test items by the two content dimensions process components and software applications

Dimensions of ICT Literacy	Operating system/ word processing 1	Spread-sheets 2	Presenta-tion/ graphics 3	E-mail/ communi-cation 4	Internet/ search engines 5	*Total*
Technological literacy						
Define	8	–	5	7	–	*20*
Access	10	2	4	4	4	*24*
Manage	3	6	2	5	2	*18*
Create	8	10	7	4	–	*29*
Information literacy						
Integrate	4	4	–	1	15	*24*
Evaluate	–	–	–	2	7	*9*
Communicate	1	5	6	4		*16*
Total	*34*	*27*	*24*	*27*	*28*	*140*

With the exception of a few items that addressed factual knowledge (e.g., computer terminology), the items asked students to accomplish computer-based tasks. To do so, students were presented with realistic problems embedded in a range of authentic situations. Most items used screenshots, for example, an internet browser, an electronic database, or a spreadsheet as prompts. As an illustration, Figure 2 presents a sample item that assesses technological literacy while Figure 3 shows an item that assesses information literacy.

Figure 2: Sample item assessing the process component *Create*

This table shows the number of tickets that have been sold for a school play. Which formula is needed in the table to calculate the total number of tickets issued on Thursday?

	A	B	C	D	E	F
1				Drama Club Play		
2				Tickets		
3		Tickets sold	Tickets issued free of charge	Total no. of tickets	Total attendance	
4	Monday	28	0	28	26	
5	Tuesday	29	24	53	53	
6	Wednesday	65	16	81	80	
7	Thursday	58	8		64	
8	*Total*	180	48	228		
9						
10						

=B7+C7	☐
=D5:D7	☐
=E8-B8	☐
=B8-C8	☐

Figure 3: Sample item assessing the process component *Integrate*

Anna wants to find out about aquariums, but <u>not</u> about aquariums using saltwater. In which of the fields below should Anna enter the words “salt water”?

Find results	related to **all** of the words	
	related to the **exact phrase**	
	related to **any** of the words	
	not related to the words	

related to all of the words	☐
related to the exact phrase	☐
related to any of the words	☐
not related to the words	☐

4. Research objectives

We investigated the psychometric quality of the newly developed item pool (Research Objective 1) as well as aspects of its validity (Research Objective 2) by addressing the following research objectives:

(1) We assessed the *psychometric quality* of the item pool by examining the reliability of the test as well as the discriminatory power and fit indices of the items. In addition, we examined the distribution of item difficulties and the ability distribution. Finally, unidimensionality of the item pool was investigated.

(2a) The *content validity* of the items constructed was examined by asking national experts to rate each item on two dimensions (each on a 4-point rating scale): (a) *How well do the abilities required to solve this task reflect the construct which is intended to measure?* and (b) *how important is the knowledge required to solve the task for students of this age group with respect to their future training and careers?*

(2b) We investigated *test fairness* by conducting DIF analyses for gender and school type. That is, we estimated task difficulties separately for subsamples of students (e.g., girls and boys) and compared the results with each other. In doing so, we aimed to identify items that compromised the fairness of the test. In addition, the DIF of the items selected for the main test was cross-validated in a second sample.

(2c) With respect to *concurrent (criterion) validity*, we expected ICT literacy to be significantly associated with various student background variables. We therefore included grade level, school track (highest German secondary school track *Gymnasium* vs. other), and the family's cultural capital as covariates in the analyses.

Furthermore, we expected that aspects of computer familiarity (e.g., whether students were motivated to use computers for entertainment; see Liaw, 2002) would make a significant contribution to explaining differences in ICT literacy. Specifically, in accordance with the functional literacy approach, we expected using computers for task-oriented purposes (e.g., looking up information) to be positively correlated with literacy scores (convergent validity), but using computers for entertainment purposes (e.g., for relaxation) not to be significantly associated with ICT literacy (divergent validity).

5. Method

5.1 Sample

The item pool was administered to a total of 855 lower secondary students in three pilot studies conducted within the context of NEPS. Table 2 reports the individual sample sizes for the three participating grade levels as well as the distribution by gender and school type. In addition, we included data from 316 Grade 5 students who were administered part of the item pool in the context of the Panel Study at the Research School "Education and Capabilities" in North Rhine-Westphalia (PARS; Bos et al., 2013). The data obtained from this sample (see also Table 2) were used for the cross-validation of the final instrument that resulted from the analyses conducted in the pilot studies.

Table 2: Overview of the study samples by grade level, gender, and school type

			Gender				School type			
	N	Grade level	Female		Male		*Gymnasium*		Other school types	
			n	(%)	*n*	(%)	*n*	(%)	*n*	(%)
K5 pilot study	231	5	124	(53.7)	107	(46.3)	95	(41.1)	136	(58.9)
K7 pilot study	316	7	168	(53.2)	148	(46.8)	134	(42.4)	182	(57.6)
K9 pilot study	308	9	157	(51.0)	151	(49.0)	118	(38.3)	190	(61.7)
Total (pilot studies)	855	–	449	(52.5)	406	(47.5)	347	(40.6)	508	(59.4)
PARS study	420	5	220	(52.4)	200	(47.6)	177	(42.1)	243	(57.9)
Total (pilot studies and PARS)	1275	–	669	(52.5)	606	(47.5)	524	(41.1)	751	(58.9)

5.2 Design and instruments

Due to the large size of the item pool (140 items), each student was administered only a subset of items. To this end, we implemented a rotated test design (multi-matrix design) in which students took different but overlapping tasks, thus ensuring that the item subsets were linked across the sample. Overall, 60 items were administered to the Grade 5 students (K5 pilot study and PARS study), 70 items to the Grade 7 students (K7 pilot study), and 80 items to the Grade 9 students (K9 pilot study). The Grade 5 and 9 students were each administered a separate set of items that did not contain overlapping anchor items, whereas the test for Grade 7 students comprised 30 Grade 5 items and 40 Grade 9 items. All items were thus linked across the four samples. That way, they could be scaled together using item-response theory and reported on the same metric.

In three of the four samples (K7 and K9 pilot studies, PARS study), we included the following variables as covariates in our analysis of how background variables influence ICT literacy: grade level, gender (0 = *female*, 1 = *male*), school type (1 = *Gymnasium*, 0 = *other*), and number of books in the home as an indicator of the family's cultural capital (5-point rating scale from 1 = *0–25 books* to 5 = *more than 500 books*).

In the K9 pilot study, we also controlled for the following variables in our assessment of criterion validity (see Table 3 for an overview showing mean scores, standard deviations, and sample items):

5.2.1 Parents' educational level

We assessed the highest educational qualification in the family, which was then converted into the number of years of education (OECD, 2007).

5.2.2 German/mathematics grades

Students' self-reported German and mathematics grades were used as indicators of their school achievement (from 1 = *insufficient/fail* to 6 = *very good*).

5.2.3 Aspects of computer familiarity

Apart from owning a computer (0 = *no*, 1 = *yes*), we assessed the frequency of computer use at home and at school on 6-point rating scales (from 1 = *never* to 6 = *[almost] every day*). Further, entertainment-related (intrinsic) versus task-oriented (extrinsic) motivations for using the computer and the internet were each assessed by six items on 4-point rating scales (from 1 = *agree* to 4 = *disagree*). Analogous scales measured use of digital media for entertainment versus task-oriented purposes.

Table 3: Variables used in the K9 pilot study to test the criterion validity of the TILT: Means (*M*), standard deviations (*SD*), reliabilities (Rel), and sample items

Variable	No. of items	*M*	*SD*	Rel[a]	Sample item
School track (Gymnasium)	1	0.38	–	–	–
Cultural capital (number of books in the home)	1	3.40	1.33	–	"About how many books are there in your home?"
Highest parental educational qualification (in years of education)	1	12.21	3.35	–	"What is the highest educational qualification held by your parents?"
German grade	1	4.19	0.86	–	"What was your German grade on your last report card?" (reverse coded)
Mathematics grade	1	4.10	1.03	–	"What was your mathematics grade on your last report card?" (reverse coded)
Own computer	1	0.67	0.47	–	"Do you have your own computer at home?"
Frequency of use at home	1	5.61	0.87	–	"How often do you use a computer at home?"
Frequency of use at school	1	2.92	1.41	–	"How often do you use a computer at school?"
Task-oriented motivations (extrinsic)	6	2.88	0.63	.82	"I use the computer to look up information on certain topics."
Use for task-oriented purposes (extrinsic)	4	3.03	0.98	.68	"I use the computer to do calculations with spreadsheet software or to create charts and graphs."
Entertainment-related motivations (intrinsic)	6	2.47	0.66	.79	"I use the internet to share my photos with my friends."
Use for entertainment purposes (intrinsic)	5	4.31	1.10	.73	"I use the computer for music, videos, and photos."

[a] Cronbach's α.

5.3 Statistical analyses

Due to the multi-matrix design, we used *ConQuest* (Wu, Adams, Wilson, & Haldane, 2007) to scale the data and to investigate the psychometric quality of the test items (e.g., to compute item difficulties, test scores, and scale reliabilities). The data were scaled using a unidimensional Rasch model, and the parameters were estimated using Marginal Maximum Likelihood (MML) techniques. Weighted Likelihood Estimates (WLE) were used as person parameters (ability estimates). The WLEs were also used to determine the reliabilities regarding both the full sample and the subpopulations (by grade level, school type, and gender) and in the DIF analyses. The analyses of criterion validity were conducted using the SPSS software package.

6. Results

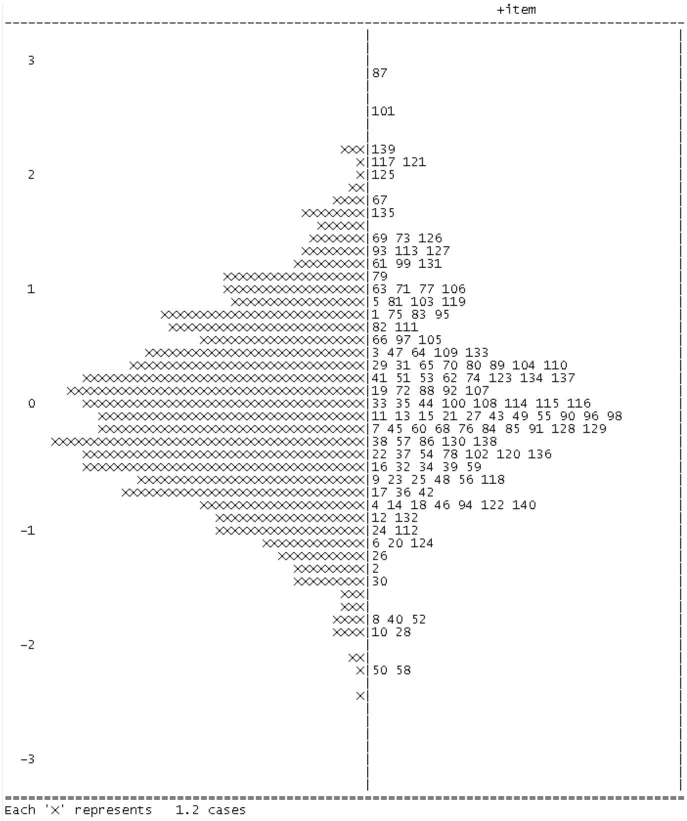
6.1 Research Question 1: Scaling and assessing the psychometric quality of TILT

The WLE reliability of the jointly scaled data from the K5, K7, and K9 samples was .91, indicating very high reliability of measurement. As shown in Table 4, the same held in the subgroups disaggregated by gender, school type, and grade level: Reliabilities for all subgroups were above .80. Moreover, reliability was clearly related to the number of items administered. For the full sample, the item discriminations (correlation between a student's score on a particular item and score on the overall test) were between -.03 and .51. The infit statistic (weighted mean square residual; MNSQ) indicates the fit of each item to the Rasch model and has the expected value of 1, with deviations of ± .15 being considered compatible with the model. The infit scores of the test items were between 0.84 and 1.25. Only six items showed a poor fit (infit scores > 1.15). These items were excluded from further analyses. The data thus indicate that the instrument provided a largely unbiased and reliable estimate of ICT literacy. As displayed in Figure 4, the comparison of the distributions of item difficulties and person abilities shows a wide spread of item difficulties providing a good match to the students' ability levels.

Table 4: WLE reliabilities in subsamples disaggregated by gender, school type, and grade level

Variable	Group	*N*	No. of items	WLE reliability
Grade level	K5	231	60	.82
	K7	316	70	.88
	K9	308	80	.90
Gender	Girls	449	140	.90
	Boys	406	140	.92
School type	Gymnasium	347	140	.89
	Other school types	508	140	.90

Figure 4: Distribution of person and item parameters in the full sample



In addition, the unidimensionality of the item pool was tested by comparing the model fit between the unidimensional model and the two-dimensional model with the factors technological literacy and information literacy. To do so, we used the *Bayes Information Criterion* (BIC) and the *Consistent Akaikes Information Criterion* (CAIC) as indicators of model fit (Rost, 2004). Both indicators showed that the unidimensional model (BIC: 75348.07; CAIC: 75489.07) showed a better model fit than the two-dimensional model (BIC: 75354.74; CAIC: 75497.74).

6.2 Research Question 2a: Assessment of content validity of TILT

The content validity of the item pool was examined by consulting national experts in ICT literacy research. With mean scores of 3.20 (Question 1: construct relevance of the abilities needed to solve the item) and 3.37 (Question 2: relevance of the abilities assessed for students' future training and careers), the results indicate that the content validity of the items was judged to be rather high (see Table

5). Detailed inspection of the expert ratings across the seven process components revealed that, on average, the technological literacy items (*Define*, *Access*, *Create*, *Manage*; M = 3.27) were rated to assess the intended construct somewhat better than the information literacy items (*Integrate*, *Evaluate*, *Communicate*; M = 3.07). In contrast, the prognostic value of the information literacy items was rated to be somewhat higher than that of the technological literacy items. This holds particularly true for the process components *Evaluate* (M = 3.75) versus *Define* (M = 3.04). Overall, however, the expert ratings showed relatively little variation across the process components and were consistently well above the scale midpoint, indicating a positive evaluation.

Table 5: Expert ratings of the items' content validity and relevance for students' future training and careers

	Relevance of the abilities needed to solve the item[a]	Relevance for training/ career[b]
Processes		
Define	3.38	3.04
Access	3.26	3.24
Create	3.33	3.45
Manage	3.05	3.46
(Technological literacy)	*(3.27)*	*(3.30)*
Integrate	2.99	3.36
Evaluate	3.61	3.75
Communicate	2.89	3.53
(Information literacy)	*(3.07)*	*(3.49)*
Software applications		
Operating system/word processing	3.40	3.34
Spreadsheets	2.91	3.36
Presentation software	3.28	3.37
E-mail/chats/forums	3.17	3.31
Internet/search engines	3.24	3.51
Total	3.20	3.37

[a]Scale: 1 = *very poorly*, 2 = *quite poorly*, 3 = *quite well*, 4 = *very well*. [b]Scale: 1 = *unimportant*, 2 = *fairly unimportant*, 3 = *fairly important*, 4 = *very important*.

Inspection of the expert data across the five software applications showed that (with the exception of spreadsheets in Question 1) the variation in expert ratings on this dimension was even smaller. In sum, the largely homogeneous and consistently positive ratings across all process components and software applications attest to the soundness of the conceptual framework.

6.3 Research Question 2b: Analysis of differential item functioning of TILT

We tested for DIF across gender and school type. Table 6 summarizes the results of these analyses. An absolute difference greater than .3 between an item's difficulty for a subgroup and its difficulty for the total sample was considered to indicate significant DIF (OECD, 2009). Of the 140 items examined, only 10 (7%) showed gender DIF exceeding this level. The proportion of items with DIF across school types was somewhat larger, at 16% (23 items). Because a main effect of ICT literacy was also found for school type, the DIF values can be understood to represent deviations from this global mean difference.

Table 6: DIF across gender and school type

		Number of items with $\lvert DIF \rvert > 0.3$	
Variable	Subgroup	K5 sample	K9 sample
Gender	Girls	0	6
	Boys	1	3
School type	Gymnasium	5	8
	Other school types	6	4

Note. N = 855. Total number of items: K5: 60, K9: 80.

6.4 Selection of items for the main test

Based on our analysis of the psychometric quality of the item pool, items were then selected for the main NEPS assessments (Grades 6 and 9). All items with significant gender DIF (> 0.3) and unsatisfactory item fit (weighted MNSQ > 1.15) were excluded. Further items were eliminated on the basis of their discriminatory power, the expert ratings, and the results of the DIF analyses across school type. Because only a test time of 30 minutes was available for each domain assessed in the main test, a set of 30 items was selected for the Grade 6 assessment, and a set of 40 items for the Grade 9 assessment.

6.5 Cross-validation

To cross-validate the results, we repeated the DIF analyses in a second sample. As this PARS sample comprised only Grade 5 students, we focused on items selected for implementation in Grade 6 in the main test. Table 7 compares the findings for the two samples. As expected, the optimized results for the K5 pilot sample (in which no items showed gender DIF) were not fully replicated in the cross-validation sample. Whereas item selection meant that none of the items in the K5 pilot study sample showed gender DIF, four items in the PARS sample exhibited gender DIF. However, the clear reduction in the number of items with DIF across school type was replicated in the validation sample. The items exhibiting DIF in the validation sample were revised for use in the main study. Hence, the main study data are not expected to show DIF across the subpopulations considered.

Table 7: DIF of the items selected for the K6 main test in the K5 sample and in the PARS sample

Variable	Group	Total no. of items	Number of items with $\|DIF\| > 0.3$	
			K5 pilot study	PARS study
Gender	Girls	30	0	1
	Boys	30	0	3
School type	Gymnasium	30	2	2
	Other school types	30	1	1

6.6 Research Question 2c: Assessing concurrent validity of TILT (whole sample)

We drew on data obtained from 1,044 students (K7 and K9 pilot studies, PARS study) to assess the influence of background variables (grade level, school type, cultural capital, and gender) on ICT literacy. Since gender was not found to be significantly related to ICT literacy ($t(1042) = -1.90$, $p > .05$), we conducted a three-factor ANOVA with the remaining variables. To this end, the *number of books in the home* variable was dichotomized (*up to 100 books* vs. *more than 100 books*).

Consistent with our hypotheses, all three background variables showed significant main effects. In terms of grade level, $F(2, 1032) = 352.75$, $p < .001$, $\eta^2 = .41$, the data revealed significant differences in ICT literacy across grades, with the higher grades outperforming the lower ones with effect sizes of $d = 1.01$ (Grade 5 vs. Grade 7) and $d = 0.81$ (Grade 7 vs. Grade 9; see Table 8 for mean scores for the individual samples by grade level). Likewise, a significant main effect was obtained for school track in favor of *Gymnasium* students, $F(1, 1032) = 305.17$, $p < .001$, $\eta^2 = .23$. The effect sizes for the individual samples were $d = 1.40$ (Grade 5), $d = 1.48$ (Grade 7), and $d = 1.15$ (Grade 9). The main effect of cultural capital (number of books in the home) was also significant, $F(1, 1032) = 18.26$, $p < .001$, $\eta^2 = .02$, with the effect of cultural capital increasing across grade levels (Grade 5: $d = 0.55$; Grade 7: $d = 0.68$; Grade 9: $d = 0.82$). Accordingly, there was a significant interaction effect between grade level and cultural capital, $F(2, 1032) = 3.08$, $p = .046$, $\eta^2 = .01$. In other words, differences in ICT literacy increased with students' grade level as a function of their family's cultural capital. None of the other interactions reached the level of significance.

Table 8: Mean ICT scores (WLE parameters) from the three-factor ANOVA by grade level

	Number of books in the home			
	≤ 100 books	> 100 books	(Total books)	Total
Grade 5				
Gymnasium	-0.38	-0.37	*-0.37*	
Other school types	-1.33	-1.11	*-1.27*	
(All school types)	*-1.06*	*-0.65*		*-0.89*
Grade 7				
Gymnasium	0.40	0.40	*0.40*	
Other school types	-0.57	-0.35	*-0.50*	
(All school types)	*-0.36*	*0.12*		*-0.12*
Grade 9				
Gymnasium	0.83	1.09	*1.01*	
Other school types	0.06	0.51	*0.20*	
(All school types)	*0.22*	*0.85*		*0.51*

6.7 Research Question 2c: Assessing concurrent validity of TILT (K9 sample)

To test further aspects of the criterion validity of the item pool, we drew on data provided by a total of 308 students (K9 pilot study) who – in addition to the test items – completed a computer familiarity questionnaire. The aspects of computer familiarity assessed were owning a computer, frequency of computer use at home and at school, task-oriented versus entertainment-related motivations for using the computer, and using computers for task-oriented versus entertainment purposes. Indicators of social background and school achievement (school type, cultural capital, highest parental educational level, German and mathematics grades) were included as covariates in the analysis. In addition to inspecting the correlation matrix, we performed sequential regression analyses to explore if the aspects of computer familiarity assessed made a significant contribution to explaining differences in ICT literacy even when indicators of social background and school achievement were controlled.

Table 9: Regression models predicting ICT literacy from social background variables and aspects of computer familiarity

	Correlation	Model 1	Model 2	Model 3
	r	β	β	β
Social background/cognitive indicators				
School type: *Gymnasium*	.48**	0.31**		0.33**
Cultural capital (no. books in the home)	.42**	0.21**		0.19**
Highest parental educational qualification (in years of education)	.27**	0.06		0.02
German grade	.29**	0.05		0.05
Mathematics grade	.41**	0.26**		0.20**
Computer familiarity				
Own computer	.15**		0.09	0.04
Frequency of computer use at home	.16**		0.15**	0.08
Frequency of computer use at school	.15**		0.16**	0.23**
Task-oriented motivations	.34**		0.37**	0.20**
Use for task-oriented purposes	.16**		-0.09	-0.09
Entertainment-related motivations	-.19**		-0.27**	-0.11*
Use for entertainment purposes	-.02		0.04	0.07
R^2		*0.38*	*0.22*	*0.47*

*p < .05. **p < .01.

Consistent with our hypotheses, ICT literacy scores correlated positively with task-oriented motivations and types of use, but negatively or not at all with entertainment-related motivations and types of use (see Table 9). The other aspects of computer familiarity showed significant relations with ICT literacy, but the size of the effects was rather modest. As expected, with 38%, the indicators of social background and school achievement explained a considerable proportion of the variance in ICT literacy scores (Model 1). The aspects of computer familiarity explained 22% of the variance in ICT literacy (Model 2). Moreover, even when the indicators of social background and school achievement were controlled, the aspects of computer familiarity explained an additional 9% of the variance in ICT literacy (Model 3). Beside frequency of computer use at school, task-oriented and entertainment-related motivations for computer use proved to be the most powerful predictors of ICT literacy.

7. Discussion

In this article, we presented the TILT, a newly developed instrument for the longitudinal assessment of technological and information literacy. Whereas previous paper-and-pencil tests have focused mainly on the technological dimension of computer literacy, our aim was to develop a test that also measures information literacy. Assessment of information literacy is of specific interest in the longitudinal context of NEPS because this aspect of ICT literacy will remain relevant despite the rapid pace of technological change (ETS, 2002). The present analyses of the instrument's psychometric quality indicate that item and scale characteristics were already satisfactory in the pilot study samples. Through the careful selection of items for the main study (e.g., exclusion of items with low discriminatory power, poor item fit, or DIF), it was possible to further enhance the psychometric quality of the TILT instrument, as shown by the results of the cross-validation. Moreover, the item pool showed a good fit to the ability level of the test takers. In terms of test fairness, DIF analyses identified items that put certain groups of students (e.g., students of one gender or students at a specific school type) at an advantage or a disadvantage, and these items were eliminated from the main studies in the respective age cohorts. The fairness of the items selected on the basis of the pilot study data was largely confirmed in a second independent sample.

Further, the expert ratings confirmed the content validity of the items developed and the soundness of the conceptual framework. On average, the technological literacy items were judged to assess the intended construct somewhat better than the information literacy items. The expert ratings indicated that more complex demands, such as using digital media to organize and structure information (i.e., the process component *Integrate*), are not optimally operationalized by multiple-choice items in paper-and-pencil format. Further studies using performance-based items administered by computer are therefore needed to test the extent to which paper-and-pencil items are capable of assessing all facets/process components of ICT literacy with sufficient validity. In contrast, the experts judged the prognostic value of the information literacy items (especially the process component *Evaluate*) to be somewhat higher than that of the technological literacy items. This evaluation reflects the perspective taken by various frameworks based on the concept of functional literacy, according to which it is the problem-oriented use of digital media that facilitates active participation in society (e.g., ETS, 2002; van Dijk, 2006). In view of the longitudinal design of NEPS, moreover, this result further highlights the importance of assessing information literacy in addition to technological skills.

Our analysis of concurrent (criterion) validity revealed the expected pattern of relations. ICT literacy showed significant associations in the expected directions with the background variables investigated (grade level, school type, cultural capital of the family). With effect sizes equivalent to approximately one standard deviation, the magnitudes of the differences across adjacent grade levels and school types were considerable and consistent with those reported in other studies (Kuhlemeier & Hemker, 2007; MCEETYA, 2007). A significant interaction effect was obtained for cultural capital: Differences in ICT literacy increased with students' grade level as a function of the family's cultural capital. Here, again, the effect size of the differences was large. The results thus point to a growing digital divide (i.e., an increasing gap in ICT literacy over time) as a function of social background. As the samples recruited for the pilot studies were independent and non-representative, longitudinal data from the main study are necessary to draw firm conclusions in this respect. Consistent with the large effect sizes, the background variables examined accounted for more than one third of the variance in ICT literacy.

Findings showing that aspects of computer familiarity (computer ownership, frequency of use, motivations for use, and types of use) explained a further 9% of the variance in ICT literacy thus attest to the criterion validity of the item pool developed. Without control for the background variables, these aspects accounted for over 20% of the variance. This result is again consistent with the findings of previous studies (e.g., Kuhlemeier & Hemker, 2007; MCEETYA, 2007). In line with the concept of functional literacy that informed the construction of the TILT, ICT literacy was positively correlated with task-oriented motivations and types of use (convergent validity), but not or even negatively correlated with entertainment-related motivations and types of use (discriminant validity).

Although the present analyses produced satisfactory results, it should be noted that key aspects of construct validity could not be addressed in this study. The main purpose of the pilot studies was to select items for the main test, thus optimizing the instrument. For time reasons, other questions concerning the validity of the instrument could not be fully addressed. With respect to the planned switch from paper-and-pencil to computer-based testing, for example, it remains to be examined if a paper-and-pencil test is able to assess computer literacy with a similar level of accuracy and validity as a performance-based test delivered by computer. Although paper-and-pencil tests have some advantages over computer-based tests (e.g., ease of administration in large-scale assessments; implementation of large numbers of items permits the broad and reliable assessment of the construct; see Kuhlemeier & Hemker, 2007), they also have a number of disadvantages. For example, paper-and-pencil tests do not assess test takers' actual ability to use computer applications, but only their theoretical knowledge of how a computer can be used to solve specific problems. Other disadvantages of pencil-and-paper tests are, for example, that tasks are less authentic and that there is no way to capture and assess students' strategies and solution processes. Therefore, to examine the construct validity of the TILT in future studies, we will split the test items into two parallel tests and present the one test in a paper-pencil format whereas we will present the other test in a computer-based format. In doing so, we are able to identify test items that are particularly valid in a computer-based format. To realize an ecologically valid test situation, only test items are used that require test takers to use existing everyday software (e.g., word processing software). Moreover, in terms of the convergent and discriminant validity of the item pool, associations with related constructs – such as reading comprehension and problem-solving ability – remain to be analyzed.

Acknowledgment

The instrument presented in this article was developed in the context of the National Educational Panel Study, which is funded by the German Federal Ministry of Education and Research (BMBF). We thank the BMBF for its support.

The following experts assessed the content validity of the TILT items: Prof. Dr. Frank Fischer, Prof. Dr. Frank Goldhammer, Prof. Dr. Matthias Nückles, Prof. Dr. Tobias Richter, Dr. Heike Schaumburg, and Prof. Dr. Katharina Scheiter.

References

Blossfeld, H.-P. (2010). Education across the life course. In German Data Forum (RatSWD) (Ed.), *Building on progress. Expanding the research infrastructure for the social, economic, and behavioral sciences* (Vol. 2, pp. 825–840). Opladen, Germany: Budrich UniPress.

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). *Education as a lifelong process. The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft [Special Issue 14].* Wiesbaden, Germany: VS.

Blossfeld, H.-P., von Maurice, J., & Schneider, T. (2011). The National Educational Panel Study: Need, main features, and research potential. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process. The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft [Special Issue 14]* (pp. 5–17). Wiesbaden, Germany: VS.

Bos, W., Stubbe, T. C., Buddeberg, M., Dohe, C., Kasper, D., Müller, S., & Walzebug, A. (2013). Framework for the Panel Study at the Research School ‚Education and Capabilities' in North Rhine-Westphalia (PARS). *Journal for Educational Research Online*. Manuscript in preparation.

Dutke, S., & Reimer, T. (2000). Evaluation of two types of online help for application software. *Journal of Computer Assisted Learning, 16*, 307–315.

Educational Testing Service (ETS). (2002). *Digital transformation. A framework for ICT literacy*. Princeton, NJ: ETS.

Fraillon, J., & Ainley, J. (2010). *The IEA International Study of Computer and Information Literacy (ICILS).* Retrieved from http://icils2013.acer.edu.au/wp-content/uploads/examples/ICILS-Detailed-Project-Description.pdf.

Hargittai, E. (2003). *How wide a web? Inequalities in accessing information online.* Princeton, NJ: Princeton University.

International Society for Technology in Education (ISTE). (1998). *National educational technology standards for students*. Washington, DC: ISTE.

Katz, I. R. (2007). Testing information literacy in digital environments: The ETS iSkills™ assessment. *Information Technology and Libraries, 26*, 3–12.

Kozma, R. B. (2009). Transforming education: Assessing and teaching 21st Century skills. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment. New Approaches to skills assessment and implications for large-scale testing* (pp. 13–23). Luxembourg, Luxembourg: European Communities.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy. *Theory into Practice, 41*, 212–218.

Kröhne, U., & Martens, T. (2011). Computer-based competence tests in the National Educational Panel Study: The challenge of mode effects. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process. The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft [Special Issue 14]* (pp. 169–186). Wiesbaden, Germany: VS.

Kuhlemeier, H., & Hemker, B. (2007). The impact of computer use at home on students' internet skills. *Computers & Education, 49*, 460–480.

Liaw, S.-S. (2002). Understanding user perceptions of world-wide web environments. *Journal of Computer Assisted Learning, 18*, 137–148.

Markauskaite, L. (2006). Towards an integrated analytic framework of information and communications technology literacy: From intended to implemented and achieved dimensions. *Information Research, 11*, 1–20.

Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem solving. *Computers in Human Behavior, 18*, 623–632.

MCEECDYA – Ministerial Council on Education, Early Childhood Development, and Youth Affairs. (2010). *National Assessment Program – Information and communication technology literacy. Years 6 and 10. Report 2008*. Melbourne, Australia: MCEECDYA.

MCEETYA – Ministerial Council on Education, Employment, Training and Youth Affairs. (2007). *National Assessment Program – Information and communication technology literacy. Years 6 and 10. Report 2005*. Melbourne, Australia: MCEETYA.

OECD – Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Science competencies for tomorrow's world: Vol. 1. Analysis*. Paris, France: OECD.

OECD – Organisation for Economic Co-operation and Development. (2009). *PISA 2006 Technical Report*. Paris, France: OECD.

Partnership for 21st Century Skills. (2007). *Learning for the 21st century*. Retrieved from http://www.p21.org/storage/documents/P21

Pinto, M., Cordon, J. A., & Diaz, R. G. (2010). Thirty years of information literacy (1977–2007): A terminological, conceptual and statistical analysis. *Journal of Librarianship and Information Science, 42*, 3–19.

Poynton, T. A. (2005). Computer literacy across the lifespan: A review with implications for educators. *Computers in Human Behavior, 21*, 861–872.

Richter, T., Naumann, J., & Horz, H. (2010). Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R) [A revised version of the Computer Literacy Inventory]. *Zeitschrift für Pädagogische Psychologie, 24*, 23–37.

Rost, J. (2004). *Testtheorie – Testkonstruktion* [Test theory – test construction]. Bern, Switzerland: Hans Huber.

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2010). Measuring internet skills. *International Journal of Human-Computer Interaction, 26*, 891–916.

van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. *Poetics, 34*, 221–235.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. H. (2011). Development of competencies across the lifespan. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process. The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft [Special Issue 14]* (pp. 67–86). Wiesbaden, Germany: VS.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *ACER ConQuest 2.0: Generalised item response modeling software*. Camberwell, Australia: ACER Press.